

## **Historic, Archive Document**

Do not assume content reflects current scientific knowledge, policies, or practices.



## **PLANT GENOME RESEARCH CONFERENCE REPORT**

Sponsored by:

**United States Department of Agriculture  
Science and Education  
December 1988**

**United States  
Department of  
Agriculture**



**National Agricultural Library**



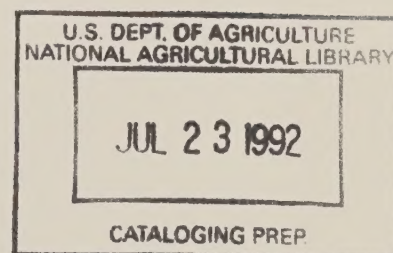
QH445  
2  
05

## PLANT GENOME RESEARCH CONFERENCE REPORT

### ACKNOWLEDGMENTS

The authors of this report are grateful to the following individuals for their assistance and support during the preparation of this report: [Faint text continues with names and acknowledgments]

Science and Education  
United States Department of Agriculture  
December 12-14, 1988



Editors:

David R. MacKenzie  
Robert M. Faust  
Keith W. Russell

PLANTING AND RESEARCH

U.S. DEPT. OF AGRICULTURE  
NATIONAL AGRICULTURAL  
LIBRARY  
MAY 3 1963  
CATALOGING

## ACKNOWLEDGEMENTS

The participants of the Crop and Forest Genome Mapping Conference held in Washington, D.C. on December 12 - 14, 1988 express their appreciation to the United States Department of Agriculture (USDA) and the Agricultural Research Institute (ARI) for sponsorship of the meeting and to the Agricultural Research Service (ARS), the Cooperative State Research Service (CSRS), and the National Agricultural Library (NAL) for organization of the meeting. Special appreciation is extended to Assistant Secretary Orville G. Bentley for his support and interest. The editorial assistance of Diane Behrens, USDA, is gratefully acknowledged.

### Conference Steering Committee

Robert M. Faust, ARS, Chairman  
Stan Cath, ARI  
David R. MacKenzie, CSRS  
Jerome P. Miksche, ARS  
Keith W. Russell, NAL





## Table of Contents

EXECUTIVE SUMMARY .....	1
Chapter I INTRODUCTION .....	4
Chapter II STATE-OF-THE-ART IN PLANT GENOME MAPPING AND SEQUENCING .....	5
Chapter III CURRENT STATUS OF EFFORTS TO STORE AND RETRIEVE GENOME INFORMATION .....	7
Chapter IV CRITERIA FOR SELECTING PLANT SPECIES .....	9
Chapter V FEATURES OF A NATIONAL NETWORK OF INFORMATION .....	12
Chapter VI ACTIVITY TO PROMOTE DEVELOPMENT AND IMPLEMENTATION .....	15
Appendix I AGENDA .....	19
Appendix II CONFERENCE PARTICIPANTS .....	25
Appendix III DEFINITIONS .....	29



## EXECUTIVE SUMMARY

Mapping the genomes of higher plants and identifying the locations of genes of agricultural importance is a high priority for agricultural research. Such knowledge will permit the manipulation of our plant genetic resources to exploit our limited natural resources more efficiently. This database will also permit the physical characterization of genes of economic importance which will renew our efforts to understand the physiological and developmental bases of variation in crops.

The goals of a USDA-sponsored plant genome mapping initiative should be:

- 1) To provide a foundation of knowledge for plant science research in the 21<sup>st</sup> century.**
- 2) To restore the preeminence of American plant genetics research.**
- 3) To improve our global position in agricultural efficiency and profitability through the deployment of varieties developed through the application of this database.**

Identifying the location of agriculturally important genes on chromosomes is now possible through the use of the tools of biotechnology. Plant breeders and geneticists have for generations sought a method by which components of agriculturally important characters could be manipulated, but have been stymied by their inability to determine which genes from each parent are found in each progeny line they test. By a process called restriction fragment length polymorphisms (RFLPs) analysis, several laboratories in the United States have demonstrated that they can now "track" each of the genes from each parent through generations to each progeny line. Through this process, in collaboration with plant breeders and statisticians, the location and value of each gene from each parent can be determined.

RFLP analysis was first applied to understanding variation in viruses, but has become an important tool in the diagnosis of genetic diseases in people. This process is complex, but technically straightforward, for all major crops. DNA is isolated from individual plants (a 2-hour process), treated with commercially available enzymes called restriction endonucleases, subjected to electrophoresis, and specific genes are identified by annealing radioactively labeled cloned pieces of plant DNA back to the electrophoresed DNA. The entire process takes 4-5 days with current technology, but improvements in technology make the process much easier, faster, and cheaper and eliminate the need for radioactive labels.

The RFLP effort in crop plants has moved along rapidly and inexpensively. Excellent RFLP maps have already been developed by U.S. scientists for corn, tomatoes, rice, lettuce, broccoli and cauliflower. Partial maps have been produced for barley, soybeans and wheat and projects are beginning for potatoes, cotton, alfalfa, dry beans, lentils, grapes and many vegetable and fruit crops.

This hybrid effort embraces the tools of fundamental molecular biology and applied crop improvement and is an example of the kind of program which must be supported if we are to help the U.S. agricultural community survive in the world marketplace. Other nations are actively pursuing plant genome mapping efforts. Japan plans to spend \$200 million specifically on efforts in rice genetics and molecular biology. Several European governments are working intensively on maps for wheat, barley, oats and vegetables. Federal support for such efforts would be expected to:

- o Enhance U.S. leadership in this research area,
- o Develop an international model for interproject database development and sharing,
- o Provide substantial basic knowledge which will enhance the U.S. competitive positions in the next century,
- o Support education and development of a cadre of trained scientists sufficient to meet current and future research needs, and



- o Enable plant breeders to design crop varieties appropriate to the stresses of a rapidly changing environment.

The conference participants developed this report to record the present "state-of-the-art" for plant genome mapping and sequencing (Chapter II) and information networks designed to distribute current knowledge and data for mapping and sequencing, (Chapter III).

Chapter II shows that the present state of knowledge of plant genome mapping and sequencing is advancing, but with irregular patterns among commodities. Reasons for these patterns include technical limitations, biological realities and funding constraints.

Chapter III reviews present applications in genome data storage and retrieval (e.g., GenBank), the Human Genome Research Project and other related research activities. Information network technology is highly advanced and extremely suitable to the needs of a national initiative in plant genome research.

The report also includes a chapter for each of three workshops conducted to develop points-to-consider in: (1) the selection of species; (2) the design of an information network; and (3) development and implementation of a national initiative in plant genome research.

When selecting species of plants for genome mapping and sequencing, the following points were identified as important:

- o Economic impact and domestic importance
- o Maximum information transfer to other plant species
- o Genome size and features
- o Yield of basic and fundamental insight
- o Existence of a good knowledge base

A national and international information network to support plant genome research should include the following characteristics:

- o User-friendly
- o Allow for all types of maps
- o Up-to-date
- o Inexpensive
- o Work with international networks

Six important items needed to initiate a plant genome research program are:

- o Clearly prioritized objectives
- o Recognition of the advantages and necessity of working with plants
- o Solicitation of some group grant proposals



- o Development of some core facilities
- o Appointment of a Plant Genome Research Coordinating Committee
- o Providing funding to support plant genome research.

The advantages of undertaking the initiative, as seen by the conference participants, were:

- o Maintenance of U.S. economic and scientific competitive positions
- o Efficient use of germplasm resources through biotechnology
- o The ability to address the fragility of natural and managed ecosystems through fundamental plant science knowledge

## Chapter I INTRODUCTION

The Plant Genome Research Conference convened in Washington, D.C. on December 12 - 14, 1988. A diverse group of plant scientists and database experts from the public and private sectors assembled to review current knowledge, identify priorities, develop guidelines, and propose plans for a major initiative to map and sequence the genomes of plant species of importance to agriculture and forestry. (The conference program is presented in Appendix I.)

The conference began with comments from Assistant Secretary Orville G. Bentley who pointed out that the USDA has a clear-cut responsibility to be concerned about plant improvement. He asked the conference participants to consider where we are today and to determine needs for an incremental plan for a national initiative on plant genome mapping and sequencing. He stated that the system should be science-based and well-planned and should incorporate an approach to information gathering and distribution.

The conference chairman, Dr. Robert M. Faust, introduced the participants and guest observers (Appendix II) and provided introductory remarks on pertinent scientific issues and initiatives.

The conference then turned its attention to the specifics of existing and needed knowledge in plant genome mapping and sequencing. The following chapters summarize the conference activities:

- o State-Of-The-Art in Plant Genome Mapping and Sequencing
- o Present Status of Efforts to Store and Retrieve Genome Information
- o Criteria for Selecting Plant Species
- o Features of a National Network of Information
- o Activity to Promote Development and Implementation

The first two chapters contain summaries of information from formal presentations made by participants in the conference. The remaining three chapters are the result of group discussions using the Nominal Group Technique (NGT) of Andre Delbecq. NGT is a highly structured group meeting procedure for identifying and organizing items related to a particular question.

## Chapter II

# STATE-OF-THE-ART IN PLANT GENOME MAPPING AND SEQUENCING

Genome mapping became feasible with the use of restriction fragment length polymorphism (RFLP) technology which allows detailed genetic mapping not possible a few years ago (see Appendix III for definitions of Genetic Mapping). Applied to plant species, this technology has created rapid advances in constructing RFLP maps that point to patterns of genomic structures and facilitate new insights into gene expression and function. RFLP technology, coupled with DNA sequencing, is expected to converge with new information in biochemistry and physiology and lead to the new understanding of how plants grow and develop and how they can be genetically engineered for improved characteristics.

However, convergence of these new technologies is not expected in the short-term. Considerable progress has been made in some crop species while others lag behind. The following brief reports highlight the current status of research on selected commodities and point out some of the advantages and disadvantages of conducting genome research on a particular species.

Maize - RFLP mapping in maize is probably the most advanced of all plant species. Five different research groups are developing RFLP maps, and about 900 markers have been mapped. Research is also being undertaken on a limited scale to unify the various RFLP maps, to rationalize the conventional and RFLP maps, and to analyze quantitative trait loci (QTL) such as for yield, maturity, disease and insect resistance. Some work is also underway on quality enhancement in sweet corn. Efficient mathematical and computer-aided procedures for mapping analysis are urgently needed for genome analysis and are being developed from research studies with maize.

Soybeans - Several research groups are currently engaged in RFLP mapping in soybeans, using both random genomic and cDNA markers. RFLP-type polymorphism is relatively low among soybean cultivars. Tissue culture has been explored as one method of inducing polymorphism, but has not proven successful thus far.

Cotton - RFLP mapping in cotton has only just begun, but it has been used for organelle DNA. Private sector research with cotton mapping RFLPs has been focused on using the technology for varietal protection through DNA fingerprinting.

Vegetables - Analysis of disease and insect resistance genes and QTL is relatively advanced in the tomato. Two groups have developed RFLP maps in the tomato, and a total of about 700 markers have been mapped. Other vegetable species receiving attention include Brassica, lettuce, potatoes, and peppers, all of which have fairly detailed RFLP maps.

Forest Species - Genome mapping in forest trees is currently restricted to conifers. There are two Forest Service projects aimed at constructing RFLP maps for loblolly pine: one at Berkeley, CA, to characterize the organization of pine genomes, and one at Gulfport, MS, to isolate fusiform rust resistance genes. Some work is also underway at several universities including North Carolina State, Kentucky and Texas A&M. Statistical methods being developed for mapping and identifying QTLs in humans will facilitate similar efforts in forest species.

Barley - RFLP mapping is fairly advanced in barley; it has allowed the location of some genes of agronomic importance and the identification of genotypes carrying alternate alleles. The North American barley breeders have coordinated a mapping effort that includes work on developing doubled haploids, map construction, field studies, cytogenetics, and analysis of QTL.

Forages - Although much of the genetic work on forages has been with grass species, forage is typically composed of both grass and legume species. Forage species tend to have complex genomes and difficult breeding systems. Breeding objectives tend to vary but include improved seed and herbage yield, quality and digestibility, and pest resistance. Most work on mapping forage species has been with alfalfa, with a number of investigators collaborating nationally. Additionally, pearl millet mapping has made considerable advances based on research being conducted by scientists in India and the United States.

Rice - Genome mapping is expected to advance rapidly in rice because of a new Japanese Government initiative to map and sequence the rice genome (\$200,000,000 over a ten-year period). An RFLP map with about 200 markers has been developed by a group in the United States, and is being used to analyze rice disease and insect resistance genes.

Wheat - British scientists have a project underway to develop an RFLP map, but the RFLP probes are proprietary. Sequence information is available through research efforts involving genes such as for ribosomal RNA, histones, chlorophyll A/B binding proteins and globulins. Chromosomal locations are known for many phenotypic and isozymic markers, and the genetics and cytogenetics of wheat are relatively advanced.



### Chapter III

## CURRENT STATUS OF EFFORTS TO STORE AND RETRIEVE GENOME INFORMATION

The conference included presentations by representatives from the National Library of Medicine, the National Agricultural Library, GenBank, and the National Institutes of Health's Office of Human Genome Research. These presentations addressed the current state of cooperative efforts related to information storage and retrieval of genome mapping and sequencing data. Highlights of these presentations are given in this chapter.

Recent legislation (Public Law 100 - 607) created a National Center for Biotechnology Information at the National Library of Medicine (NLM) and provided funding of \$8 million for fiscal year 1989. NLM will not necessarily be the physical center for all biotechnology information activities, but will support national needs through a variety of activities, including sponsoring conferences and workshops, linking information networks, and carrying out other planned projects.

NLM is developing a Directory of Biotechnology Information/Resources, a centralized computerized directory of international sources of publicly available biotechnology information. NLM is also the lead agency sponsoring a series of interagency meetings of a working group on a Biotechnology Environmental Release Databank. The group is investigating the need for a database containing information on releases into the environment of biotechnology-altered organisms.

The Library has developed a prototype computer workstation for genetics researchers. Called GenInfo, the workstation includes access to a range of computerized sources of information, including GenBank, EMBL, the Protein Information Resource, X-ray crystallographic data, related bibliographic files, and an on-line copy of McKusick's Mendelian Inheritance in Man. Several investigators at the National Institutes of Health are currently testing the system.

NLM has issued a request for applications in the area of molecular biology related to representation and analysis of molecular biology data by computer. Over the next year, NLM will also sponsor a series of small conferences on the E. coli genome.

NLM is working closely with the National Agricultural Library (NAL) to ensure that between the two libraries all biotechnology information is acquired and indexed. They are also making sure that books and articles containing genome data are indexed so they can readily be retrieved from MEDLINE and AGRICOLA (these are key NLM and NAL databases), and that information is conveyed to GenBank in a timely and useful fashion.

NAL, which serves as both the library of the USDA as well as the national library for agriculture, is involved with NLM and other agencies in efforts to make biotechnology information readily available to USDA staff and the entire international research community. A Biotechnology Information Center was established in 1986 to focus on the provision of biotechnology resources, reference services, special publications, and other needed services on this important topic. Particular emphasis has been placed on expanding the coverage of biotechnology information in AGRICOLA.

GenBank, the Genetic Sequence Databank, collects nucleotide sequence data and makes those data widely available to the international scientific community. The effort provides a computer database of all published DNA and RNA sequences and related bibliographic and biological information. Unpublished data are also included. Currently, RFLP map information is not included. GenBank is funded through a National Institute of General Medical Sciences contract with IntelliGenetics, Inc., which oversees the distribution of data and contracts with the Department of Energy to have the data collection and data storage work done at Los Alamos National Laboratory. Data collection is carried out in collaboration with the EMBL Data Library and the DNA Bank of Japan. Data are available in several machine-readable formats, both through GenBank and several secondary sources. GenBank currently contains approximately 21,000 entries covering about 25,000,000 nucleotides from all taxa of organisms.

Recent changes in the focus of GenBank include efforts to expand the database, decentralize the collection of data, and gather data from a wider range of sources. Strong encouragement is given for the submission of data by all researchers. Changes have been made to make submission of data as easy as possible and ensure its timely entry. Many journals now require the submission of sequence data to GenBank if such data are included in the paper. USDA, NIH, and other agencies can help by communicating to their researchers and grant recipients the importance of submitting nucleotide sequence data to GenBank. The ultimate goal is to make GenBank data and services better and more available to the scientific community.

The Office of Human Genomic Research at NIH will provide planning, coordination, and leadership for the NIH initiative to sequence the human genome (3 billion base-pairs). The Office presently has an outline of the project, but all of the technology is admittedly not in place. Sequencing DNA is a slow and cumbersome process which could best be done by machine if the technology can be developed. The plan is to use RFLP maps and overlapping clones to develop a physical map.

The program will become a very large resource of information and materials that will be available to the whole biological research community. The main goal is human medicine, but the impact of the project goes much further. The technology is expected to be applicable far beyond humans and will certainly impact all of biotechnology.

An NIH Ad Hoc Advisory Committee<sup>1</sup> has recommended the following to the Office of Human Genome Research:

- o In addition to humans, select model organisms with small genomes and assure that comparisons with other organisms will be possible.
- o Work on developing better analytical technology.
- o Do mapping and sequencing simultaneously.
- o Use investigator-initiated research (not research contracts), to result in better ideas and better research.
- o Make sharing of data and genetic materials, and cooperation between several agencies, organizations and laboratories, a cornerstone of this program.

The Ad Hoc Human Genome Project Advisory Committee also recognized the importance of resources to support the research. It is anticipated that up to \$200 million per year for a ten to fifteen-year period should be made available on a scaled-up funding system.

The International Fund for Agricultural Research is organizing a "blue ribbon" set of participants in a network for a maize genome project. As a separate program, the Department of Energy's Human Genome Program, focuses on automated technology development that will make the human genome task more manageable, but also will provide tools for other species, including plants.

---

<sup>1</sup> Other human genome study reports have been prepared by 1) The National Research Council, 2) The U.S. Office of Technology Assessments and 3) A U.S. Department of Energy Health and Environmental Research Advisory Committee. The points cited above are consistent with these additional reports.



## Chapter IV

### CRITERIA FOR SELECTING PLANT SPECIES

Conference participants identified 56 considerations related to the question: "What are the criteria for selecting crop and forest species for genome mapping and sequencing?" Those 56 considerations were evaluated and then relationally associated as:

1. Economic Impact and Domestic Impact
2. Maximum Information Transfer to Other Crop Plants
3. Genome Size and Features
4. Yield of Basic and Fundamental Insights
5. Existence of a Good Knowledge Base

In selecting plant species for genome research the likely economic benefit was the most important priority for conference participants. Related to this area was the domestic importance of the plant species selected for study and several other relevant items:

- o Technology - How will the technology be transferred and what will be the short-term benefits of a plant genome mapping and sequencing initiative? Consideration should be given to avoiding such conflicts as "big science" versus "small science," and to determine how the technology will impact foreign markets, less developed countries, and world agriculture, especially given the opportunities for international collaboration. The selection of plant species for regional importance may become an essential point to consider.
- o Finances - The conference participants raised the question of the availability of resources to support a major initiative in plant genome mapping and sequencing. To obtain full value for the money invested, great care should be taken to avoid duplication of research effort and to encourage careful use of resources.
- o Nutrition - It is very likely that some of the practical outcomes of a plant genome research initiative will greatly improve the nutritional value of harvested plant species. This point should be considered in planning the program.
- o Proprietary Information - Careful thought should be given to intellectual property rights, availability of knowledge, and sharing of that knowledge within the scientific community. Openness in exchanging information should be actively promoted.
- o Potential for Ecological Impact - Consideration should be given to research that will work to improve or protect the environment.

The conference participants identified as a second priority that the species selected for research should provide maximum information transfer to other crop plants. This consideration emphasizes the need to use resources wisely for discoveries which have broad application rather than being limited to specific taxonomic groups. The selected species should serve as models within species groups and should collectively represent a variety of breeding systems. Consideration should also be given to the pool of available investigators and the genetic and cytogenetic stocks available for research. The ease of access to knowledge from quantitative studies as well as the potential for developing fundamental insights are important. The research should maximally contribute to developing basic knowledge and to providing some scientific "leverage" to extend the knowledge beyond simply the direct applications.

The third high priority identified by the conference participants was the suitability of a species for genomic analysis. Consideration should be given to the amenability of the genome to analysis and to the size and

complexity of the genome. Care should be taken to select species that will provide timely progress in developing the technology, while at the same time providing some taxonomic diversity in species studied. The species selected should allow the application of existing techniques. The degree of polymorphism should also be considered along with the number and availability of genetic and cytogenetic stocks, the extent of existing genetic maps, and the suitability for QTL study. Consideration should be given to genomic stability, generation time, ability to do gene tagging, and the degree of knowledge about genome organization. Finally, the extent and status of germplasm collections of the species should be evaluated.

In addition to the above, other criteria for selecting plant species for genome mapping and sequencing should include the likelihood of developing basic knowledge and fundamental insights in basic biology. Some points to consider in this aspect include:

- o Extent of existing knowledge.
- o Degree of taxonomic diversity.
- o Representative breeding systems.
- o Extent of pre-existing genetic maps.
- o Existing quantitative studies.
- o Existing genetic and cytogenetic stocks.
- o Scientific leverage (i.e., information transfer to other species).
- o Exploitation and complementation of existing knowledge bases in molecular biology, biochemistry, and physiology.
- o Knowledge of genome organization.
- o Recognition of interesting target genes.

The final major priority identified by the conference participants was that plant species should be selected that have a good knowledge base. For obvious reasons careful selection should lead to the fastest and easiest progress of the research. This knowledge base might be preexisting genetic maps, genetic and cytogenetic stocks, and germplasm collections. Some points to consider in selecting plant species based on the existence of good knowledge bases are:

- o The ability to transfer existing technology.
- o The possibility for short-term impact.
- o Existing knowledge in biochemistry and physiology.
- o Status of current, similar projects.
- o Predicted probability of success.
- o Track record of information exchange.

One of the expected benefits from this approach is the contribution to the development of basic knowledge in this area. The research should also develop the potential for fundamental insights for other activities in research.

Points that should be considered when selecting plant species for genome mapping and sequencing include the ease of large-scale transformation (molecular gene transfers) and the ability to regenerate whole plants from single cells. The conference participants also addressed the question of how the final selection of



crops should be made; as commodities or on identified needs for technology development? Moreover, there is a risk of a "big science" versus "small science" confrontation by focusing resources on a few species and the possibility of potential duplication of efforts in the private sector. Hence, some degree of coordination, with as free exchange of information as possible, will be required.

## Chapter V

# FEATURES OF A NATIONAL NETWORK OF INFORMATION

In the second working session, the participants of the conference addressed the question: "What are the features of a national information network of plant genome data?" Seventy-two items were generated, sorted and re-evaluated, resulting in four major recommendations:

1. The network should be user-friendly with systematic organization for input.
2. The network should allow for all types of maps, quantitative information and raw data.
3. The network should be kept current with frequent updates, and include a mechanism for data validation.
4. Use of the network should be free or relatively inexpensive.

The design of the information system is critically important to the network's success. It is more likely to be used if it is user-friendly and has a logical and systematic process of data input. Other design considerations suggested include:

- o Reaching early agreement on key words and other standardization.
- o Emphasizing a uniform, ordered way to input, organize, and relate data.
- o Providing editing capability.
- o Allowing for data quality control.
- o Including data from foreign nations.
- o Setting requirements for data validation.
- o Providing background information on crop plants (e.g., genome size).
- o Making available the sources of maps and sequence data.
- o Providing easy/unrestricted access.
- o Developing a mechanism to avoid duplication.
- o Appointing an advisory committee to aid in establishing rules.
- o Developing a procedure to collect data in an orderly fashion. (This might include the requirements for submission of data by all researchers.)
- o Providing for documentation of entries.
- o Supplying gene symbol identification.

The conference participants made several recommendations concerning the organization and structure of the system. The network should have a well-defined and structured organization, yet be flexible in order to accommodate future needs and technological advances. The database should be relational in structure to provide access to related information. It should handle different types of computer formats and be capable of providing periodic evaluation of its resources.

Conference participants envisioned overall coordination of a network of separate databases, probably at several sites, using the same software, and with standardized features. The network should be built on existing networks, to be compatible with them and yet not duplicate them. It should use standard, commercially available software. Consideration should be given to developing graphic displays of information at the user interface.

The system should "gateway" through and/or be interactive with GenBank and other established, recognized database sources (possibly including GRIN). The National Library of Medicine and the National Agricultural Library should be included for related activities.

The source code should be transparent to the user and be capable of handling different computer formats. The system should be available on-line (and in other formats), with good indexing and search options and downloading capability. It should have the ability to cross reference between maps and available probes, analyze linkage data, and permit experimental modeling.

The system structure must be developed to support a variety of complex mapping approaches. It should focus on providing "soft-copy" (not paper) and could serve a valuable function by including a bulletin board (with conferencing ability) on topical information.

Inasmuch as this national information network for plant genome data would predominantly be a service to the research community, the mixture of types of information to be stored for retrieval is complex. The demands on the system will be extensive and dynamic. Inclusion of different types of maps, some requiring co-presentation for analysis, will require sophisticated approaches. Inclusion of raw data is seen as critical for some types of analytical processes and should be given strong consideration. The availability of both maps and sequence data in one single system is a very important feature. References to related, published information must also be available in the system.

Information will need to be stored that describes the identified function of specific sequences and can cross-identify identical sequences in different plant species. The combination of different types of maps (e.g., phenotypic and RFLP) in one system will be especially useful to researchers.

There will be a need to allow the analysis of data (e.g., linkage information) and permit experimental modeling with archived and newly entered data. There should be a roster of researchers identified by different types of categories, and catalogs of germplasm available to researchers.

Information should be provided on synteny (i.e., common order of information along chromosomes in different species) and provisions made to allow (indeed encourage) submissions of partial DNA sequences to accompany RFLP data. The system should also accommodate organelle DNA information including cytoplasmic and mitochondrial functions. There should be descriptions of probes that include the enzyme used and fragment size.

Currency and validity are two important coordination characteristics of useful information systems. One way of providing for these needs is the appointment of a Research Coordinating Committee to continually oversee the information network and to recommend rules for participation in the system. Quality control of data input will be a critical, ongoing concern. There must be a mechanism established for data validation, including the documentation of entries.

One way of maintaining the currency of information is to provide listings of related publications and, in the opposite direction, trigger refereed journals to provide information to the network. Consideration should be given to strongly encourage principal investigators to submit data to the system in the same way as is currently done for Chemical Abstracts and other services.

Mechanisms will have to be developed to avoid redundancy of information within the system and duplication with other networks.

In addition to being user-friendly, inexpensive or free access to the system would contribute to its success.

Some other topics not directly related to the above items were also identified:

- o An extended aspect of the information network might be the establishment of physical resources to store genetic clones as they relate to the information network.
- o The system should be designed to attract multiple funding to support the service.
- o Some mechanism should be developed to handle proprietary information.
- o The information in the system should somehow identify data as public or proprietary, including patent information on not only the sequence, but also the intended commercial uses of the discovery.
- o The system should be secure to prevent unauthorized modification of data.



## Chapter VI

### ACTIVITY TO PROMOTE DEVELOPMENT AND IMPLEMENTATION

In its third working session, the conference participants addressed the question: "What activities should the USDA undertake to promote the development of a national initiative on plant genome research?" The participants generated seventy-five items addressing this question. These items were then grouped and associated into six categories:

1. The USDA should develop clear objectives of the problem and establish the priorities.
2. The advantages of working with plant genomes and the added value of research on plants should be clearly spelled out.
3. The USDA should accept and fund group proposals for mapping model species.
4. Some core facilities should be established for plant genome studies.
5. The USDA should appoint a comprehensive Research Coordinating Committee on plant genome research.
6. Funding to support activities on plant genome research should be made available.

The first topic identified the need to more clearly define what is meant by plant genome research, including mapping and sequencing. The objectives of such research activities should be spelled out and phrased in a form understandable beyond the scientific community. An evaluation of the true magnitude of the effort should be conducted and the different elements of the task should be identified. The priorities for genome mapping and sequencing should be developed and the total information "package" should be specified.

Consideration should be given to appointing an ad hoc Research Coordinating Committee to design an organizational structure and, perhaps, even appoint a Director of an Office of Plant Genome Research (OPGR). The establishment of an OPGR will be important. Activities will need to be coordinated with other projects, such as the human genome effort. Choices will need to be made at an early date, such as which specific technology should be developed or which individual crops should be selected. There will be a need to target certain information technology and transfer methods. Thus, a Research Coordinating Committee could be very helpful in setting this process in motion.

Finally, the early identification of someone to champion the cause could be very instrumental in getting the initiative under way.

The second item identified by the group was the need to develop a clear statement on the advantages of conducting research on plant genomes. The biological advantages of conducting research with plants are their exceptional characteristics of photosynthesis, photoperiodism and special synthetic and degradative pathways. These include the synthesis of essential amino acids and the glyoxylate shunt for the conversion of fatty acids to sucrose. These are all systems which are restricted to plants.

Other biological advantages of conducting research with plants include unique hormones, the presence of apical and lateral meristems, the transduction of environmental signals into physiological and morphological responses and the synthesis of complex secondary products, often having great commercial value. Finally, plants offer the important research advantage of permitting the regeneration of whole organisms from individual cells.

Some of the genetic advantages of conducting research with plants include their reproductive systems which allow the establishment of unique individual lines and populations, possibilities for large population sizes, ease of introducing heritable traits, and analytical procedures for complex genetic inheritance.

There would seem to be a need to develop examples of the value and uses of conducting research with plants in order to better promote the benefits with individuals beyond the plant science community. Some of this information might be obtained informally from the scientific community, or as part of an organized international meeting on plant genome research. Another approach might be to publish a book on the state-of-the-art of plant genome research and the expected applications and projected future directions. Another suggestion would be to distribute this conference report and other information broadly to congressmen, lobbyists, client groups, and others.

Another approach might be to address the consequences of not undertaking the research. These consequences include a diminished competitive position in U.S. agricultural products in world markets and a lessened U.S. competitive position for agricultural science. These "costs of lost opportunities" could be evaluated as projected impacts.

The third topic identified by the group dealt with organizational aspects of the research. It was proposed that the USDA consider soliciting proposals from groups of scientists for mapping model species. The mechanism for undertaking this activity has not been specified but might be set up as new areas in competitive grants, as regional research, or as consortia for special funding.

Some other activities that would work towards early establishment of a national initiative in plant genome research include:

- o Promoting training in quantitative genetics and breeding, both at the graduate level and for more advanced scientists. [Note: Plant breeders would benefit from increased knowledge in the new methods, and scientists working with the new methods could benefit from more experiences in quantitative genetics and breeding.]
- o A demonstration project should be established to draw attention to the needs and to "show the way" in plant genome research.
- o As a first step, one species should be selected to lay out the costs and a timetable for sequencing an entire plant genome. This information might then serve as a basis for evaluating the true dimensions of a full national initiative on many crops.
- o Research should be undertaken to demonstrate the relationships between genetic information and plant physiology and biochemistry.
- o An interim project of high priority for plant genome mapping should be established some time soon within the competitive grants program.
- o A focus should be placed upon developing suitable information technology and information transfer techniques.
- o Attention should be given to mapping whole genomes and not just individual genes.

The need for funds for curating biological collections and for supporting postdoctoral associates was noted. The need to de-emphasize routine laboratory work and stress the development of new technologies was also seen as important. Finally, there was a recognized need to document present ongoing activities and to judge the capacity of current personnel to undertake a major national initiative in plant genome research.

The fourth item identified during this working group session was the need to establish core facilities for plant genome studies. Some thought should be given to the development of data collection centers that will be either 1) based at a central point, or 2) somehow distributed by crops. Decisions will be needed on how to establish priorities for such core facilities and how to identify the resources and organization that would support the effort. There may be a need for a central coordinating office for facilities and perhaps even the appointment of a core facilities director, which would then work toward better coordination of facilities to support a national initiative.



The conference participants recognized the need for USDA to accept the responsibility for being the lead agency for plant genome database networking. Procedures should be established for protocols and procedures to distribute probes within the scientific community and to provide a repository for master genetic stocks. Computer specialists should be identified to work with plant biologists in a planned and coordinated way to develop needed computer hardware and software.

Finally, an immense amount of effort will be required to make certain that other agencies, and other non-plant efforts (especially those of the NIH Office of Human Genome Research), and other public, private, and international projects are coordinated with, and complementary to, the USDA initiative.

The next item identified by the conference participants was the need to appoint a Research Coordinating Committee for a national initiative on plant genome research. The coordinating committee should help establish priorities and enlist the endorsement of high visibility scientists, as well as other agencies and institutions. It should help identify the resources needed to conduct the research. The coordinating committee should also help with central coordination, interagency coordination, and with others working in plant genome research. The coordinating committee should be looked to for advice on how funds should be distributed within the scientific community. It should suggest mechanisms for program implementation and evaluate the success of the program. The coordinating committee should also assist in obtaining support from agricultural industry, the university community and scientific societies. It should work to promote the contributions of plant genome research for enhancing national pride and prestige, both within the scientific community and with the public-at-large. It should assist in addressing the need to document ongoing activities and evaluate the capacity of current personnel to undertake aspects of the initiative. Finally, the committee should help coordinate activities for the scientific community and establish the various rules, requirements and/or guidelines for data dissemination to information networks such as GenBank.

The dimensions of a national initiative on plant genome research may ultimately be restricted by financial constraints. Broad-based support for the initiative will be critical for not only obtaining Federal funding, but also for securing matching funds from the private sector. The extent of funding will determine the degree of success as well as the amount of international participation. These aspects are considered critical for a successful national initiative in plant genome mapping.

It was proposed that the USDA consider having an NIH-style training grant program in plant genome research.

A realistic budget estimate should be developed for a major initiative in plant genome research. Consideration should be given to the full dimensions of a research program in the long term, and the desirability of providing continued funding for projects for extended periods. Thought should also be given to coordination with the National Initiative for Agricultural Research currently under consideration in the Congress - perhaps making this project part of that effort.

The economic benefits of the proposed research need to be developed more clearly and effectively in order to win broad support for an initiative.

To communicate the program needs, the conference participants recommended that the USDA sponsor a symposium on plant mapping and sequencing plant genomes. This activity, or other such activities, should publicize broadly the accomplishments, plans, and expected benefits of an initiative.

It was a strongly held view of the conference participants that the USDA's Competitive Grants Program should not simply be redirected into plant genome research, but should be supplemented with new dollars (or funds from other sources) to complement the existing competitive research grant activities.

The conference participants gave some attention to specific activities beyond those of science that would help promote the initiative in plant genome research. It was suggested that an effort should be made to "sell" the idea to Congress inasmuch as significant funding will be required to sustain the activity. Some thought should be given to developing a legislative mandate, much like the human genome research effort has done with biotechnology. There will be a need to communicate the program's benefits to commodity groups, crop improvement associations, other client groups, and the public at large.



It was proposed that the USDA evaluate the possibilities of establishing soon a prototype grants and training program to support projects in plant genome research.

A final point was made for a central focus of ideas and authority for the initiative. This authority might be achieved through the appointment of a Research Coordinating Committee, the establishment of an Office of Plant Genome Research, or some other mechanism to provide the needed leadership and energy to move an initiative forward.

## Appendix I

APPENDIX I

CROP AND FOREST GENOME MAPPING CONFERENCE

Cooperative State Research Service, USDA  
Room 338 Aerospace Building, 901 D Street, S.W.  
Washington, DC

December 12-14, 1988

MONDAY, DECEMBER 12  
A.M.

8:30

WELCOME

Dr. Orville Bentley  
Assistant Secretary for Science and Education  
U. S. Department of Agriculture  
Washington, DC

Mr. Stan Cath  
Executive Director  
Agricultural Research Institute  
Bethesda, MD

COMMENTS BY CHAIRPERSON

Dr. Robert Faust  
Agricultural Research Service  
U.S. Department of Agriculture  
Beltsville, MD

9:00

WHERE ARE WE TODAY ON PLANT GENE SEQUENCING?

OVERVIEW:

9:30 - 9:45 CORN - Dr. D. Hoisington  
University of Missouri

9:45 - 10:00 SOYBEANS - Dr. B.F. Matthews  
Agricultural Research Service (ARS), USDA

10:00 - 10:15 COTTON - Dr. R. J. Kohel  
Texas A&M University

10:15 - 10:30 VEGETABLES - Dr. James Nienhuis  
Native Plants Incorporated



10:30 - 10:45 FOREST - Dr. D. Neale  
Pacific Southwest Forest and Range Experiment  
Station

10:45 - 11:00 BARLEY - Dr. T. K. Blake  
Montana State University

11:00 - 11:15 FORAGES - Dr. David Sleper  
University of Missouri

11:15 - 11:30 RICE - Dr. R. L. Rodriguez  
University of California, Davis

11:30 - 11:45 WHEAT - Dr. Frank Greene  
ARS, c/o National Science Foundation

LUNCH - "ON YOUR OWN"

MONDAY, DECEMBER 12

P.M.

1:00 - 1:30 NOMINAL GROUP TECHNIQUE WORKING SESSIONS

Dr. David R. MacKenzie  
Cooperative State Research Service (CSRS), USDA

1:30 - 4:30 WORKSHOP #1: WHAT ARE THE CRITERIA UPON WHICH THE USDA  
SHOULD CONSIDER WHEN SELECTING CROP AND FOREST SPECIES  
FOR GENOME MAPPING AND SEQUENCING?

Co-moderators:

Dr. David Sleper  
University of Missouri

Dr. David MacKenzie  
CSRS, USDA

4:30 - 5:30 SESSION SUMMARY

6:00 - 7:30 RECEPTION AND MIXER - LEWIS ROOM, HOLIDAY INN CAPITOL

DINNER - "ON YOUR OWN"

TUESDAY, DECEMBER 13

A.M.

8:30 - 12:00 CURRENT STATUS OF COOPERATIVE EFFORTS TO COLLECT,  
ARCHIVE AND DISSEMINATE GENOME DATA

Moderator:

Mr. Keith Russell  
National Agricultural Library

Discussant: Dr. James J. Ferguson  
Specialized Information Services  
National Library of Medicine

Discussant: Mr. Jamie Hayden  
Genbank  
Los Alamos National Laboratory

Discussant: Dr. Elke Jordan  
Office of Human Genome Research  
National Institutes of Health

LUNCH - "ON YOUR OWN"

TUESDAY, DECEMBER 13

P.M.

1:30 - 4:30 WORKSHOP #2: WHAT ARE THE FEATURES OF A NATIONAL  
NETWORK OF GENE MAPS?

Co-moderators:

Mr. Keith Russell  
National Agricultural Library

Dr. David MacKenzie  
CSRS, USDA

4:30 - 5:00 Summary of Session

WEDNESDAY, DECEMBER 14

A.M.

8:30 - 11:30    WORKSHOP #3:    WHAT ACTIVITIES SHOULD THE USDA UNDERTAKE  
TO PROMOTE THE DEVELOPMENT AND IMPLEMENTATION OF A  
NATIONAL INITIATIVE IN PLANT GENOME RESEARCH?

Co-moderators:

Dr. J. Miksche  
ARS, USDA

Dr. David MacKenzie  
CSRS, USDA

11:30 - 12:00    Summary of Session

WEDNESDAY, DECEMBER 14

P.M.

Session moderators will prepare final draft of  
session summaries for inclusion in report to Science  
and Education, USDA.



## Appendix II

## Appendix II

### CONFERENCE PARTICIPANTS

**ABBOTT, Dr. A. G.**  
Dept. of Biological Sciences  
132 Long Hall  
Clemson  
Clemson, SC 29634

**BEHRENS, Ms. Diane C.**  
ARS-USDA  
212-W, Administration Building  
Washington, DC 20250

**BLAKE, Dr. T. K.**  
Plant and Soil Science  
Department  
Montana State  
Bozeman, MT 59717

**CATH, Mr. Stan**  
Agricultural Research  
Institute  
9650 Rockville Pike  
Bethesda, MD 20814

**COE, JR., Dr. E. H.**  
ARS-USDA  
Department of Agronomy  
210 Curtis Hall  
University of Missouri  
Columbia, MO 65211

**FAUST, Dr. Robert M.**  
ARS-USDA  
National Program Staff  
Building 005, Room 236  
BARC-West  
Beltsville, MD 20705

**FERGUSON, Dr. James J.**  
Specialized Information  
Services  
Bldg. 38-A, Room 35-317  
National Institutes of Health  
National Library of Medicine  
8600 Rockville Pike  
Bethesda, MD 20894

**FINCHER, Dr. Robert**  
Department of Biotechnology  
Research  
Pioneer Hi-Bred International  
Inc.  
7250 NW 62nd Avenue  
Johnston, IA 50131-0038

**GREENE, Dr. Frank**  
Developmental Biology,  
Division of Cellular Biosciences  
National Science Foundation  
1800 G Street, N.W., Room 321  
Washington, DC 20550

**GREENING, Dr. Peter**  
International Fund for Agricultural  
Research  
c/o Winrock International  
Rosslyn Plaza  
1611 North Kent Street  
Arlington, VA 22209

**HANNAH, Dr. L. C.**  
Interdisciplinary Center for  
Biotechnology Research  
Agricultural, Medical & Life  
Sciences  
1301 Fifield Hall  
University of Florida  
Gainesville, FL 32611

**HAYDEN, Dr. Jamie**  
GenBank  
T-10, MSK 710  
Los Alamos National Laboratory  
Los Alamos, NM 87545

**HOISINGTON, Dr. David A.**  
Department of Agronomy  
University of Missouri  
Columbia, MO 65211

**JORDAN, Dr. Elke**  
Director, Office of Human  
Genome Research  
Building 1, Room 332  
National Institutes of Health  
Bethesda, MD 20892

**KOCHERT, Dr. Gary**  
Department of Botany  
University of Georgia  
Athens, GA 30602

**KOHEL, Dr. Russell J.**  
Southern Crops Research  
Laboratory  
ARS-USDA  
P.O. Drawer DN  
College Station, TX 77841

**KRUGMAN, Dr. Stanley L.**  
U.S. Forest Service-USDA  
Rosslyn Plaza Building, E.  
Room 1205  
Rosslyn, VA 22209

**LARSON, Ms. Jean**  
Biotechnology Information  
Center, Room 301  
National Agricultural Library  
Beltsville, MD 20705

**LOWER, Dr. R. L.**  
College of Agriculture and  
Life Sciences  
University of Wisconsin  
Madison, WI 53706

**MacKENZIE, Dr. David R.**  
CSRS-USDA  
National Biological Impact  
Assessment Program  
Room 330L Aerospace Building  
901 D Street, S.W.  
Washington, DC 20251-2200

**MATTHEWS, Dr. Benjamin**  
Plant Molecular Biology  
Laboratory  
ARS-USDA  
Building 006, BARC-West  
Beltsville, MD 20705

**MIKSCHE, Dr. Jerome P.**  
ARS-USDA  
Building 005, Room 233  
BARC-West  
Beltsville, MD 20705

**NANIS, Ms. Suzanne**  
Biotechnology Information  
Center, Room 301  
National Agricultural Library  
Beltsville, MD 20705

**NEALE, Dr. David B.**  
Institute of Forest Genetics  
Pacific Southwest Forest and  
Range Experiment Station  
Forest Service, USDA  
Berkeley, CA 94701

**NIENHUIS, Dr. James**  
Native Plants Incorporated  
417 Wakara Way  
Salt Lake City, UT 84108

**PARRY, Dr. Richard**  
ARS-USDA  
Building 005, Room 401  
BARC-West  
Beltsville, MD 20705

**PHILLIPS, Dr. R. L.**  
Department of Agronomy and  
Plant Genetics  
University of Minnesota  
St. Paul, MN 55108

**RABSON, Dr. Robert**  
Division of Energy Bioscience  
Office of Basic Energy Sciences, ER-17  
Department of Energy  
Washington, DC 20545

**RODRIGUEZ, Dr. Raymond L.**  
Department of Genetics  
University of California  
Davis, CA 95616

**RUSSELL, Mr. Keith W.**  
National Agricultural Library  
U.S. Department of Agriculture  
Beltsville, MD 20705

**SELZER, Dr. Gerald**  
National Science Foundation  
1800 G Street, N.W., Room 312  
Washington, DC 20550

**SISCO, Dr. Paul H.**  
ARS-USDA  
Department of Crop Science  
Box 7620  
North Carolina State University  
Raleigh, NC 27695-7620

**SLEPER, Dr. David A.**  
Department of Agronomy  
209 Waters Hall  
University of Missouri  
Columbia, MO 65211

**SPECHT, Dr. J. E.**  
Department of Agronomy  
University of Nebraska  
Lincoln, NE 68583-0915

**TANKSLEY, Dr. S. D.**  
Department of Plant Breeding  
and Biometry  
Cornell University  
Ithaca, NY 14853



**TINGEY, Dr. Scott**  
E. I. du Pont de Nemours &  
Co., Inc.  
Agricultural Products Dept.  
Experimental Station  
P.O. Box 80402  
Wilmington, DE 19880-0402

**VAN BUIJTENEN, Dr. J. P.**  
Forest Genetics Laboratory  
Texas A&M University  
College Station, TX 77843-2131

### Appendix III

## Appendix III DEFINITIONS

### I. Recombinational Maps.

- o A map defining linkage relationships and meiotic recombination distances (in map units or centimorgans) between markers.
  - A. Conventional Maps.  
Maps based on morphological, physiological or biochemical markers.
  - B. RFLP Maps.  
Maps based on the use of cloned DNA fragments as markers.

### II. Physical Maps.

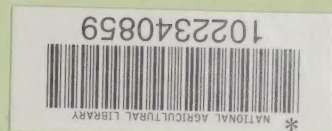
- o A map in which the distance between markers is expressed/defined in physical units (kilobase pairs, relative distance along meiotic/mitotic chromosomes).
  - A. Cytogenetic Maps.  
Maps where the relative position of markers is expressed as a distance along either meiotic (pachytene) or mitotic chromosomes.
  - B. Restriction Maps.  
Maps where the distance between markers is expressed in terms of base pairs. Distances determined by one or more of the following methods: Pulse field gel electrophoresis, restriction mapping, physical linkage.
  - C. Contig Maps  
Maps which describe a series of overlapping genomic clones (cosmid, YAC, etc.)

### III. DNA Sequence

- o The primary nucleotide sequence of DNA.







an



1022340859